Chin. Sci. Bull. (2014) 59(26):3352–3361
DOI 10.1007/s11434-014-0480-6

csb.scichina.com
www.springer.com/scp

Article

Information Processing

The Chinese "Human Flesh" Web: the first decade and beyond

Fei-Yue Wang · Daniel Zeng · Qingpeng Zhang · James A. Hendler · Jianping Cao

Received: 14 January 2014 / Accepted: 17 March 2014 / Published online: 18 June 2014

Abstract Human flesh search (HFS), a Web-enabled crowdsourcing phenomenon, originated in China a decade ago. In this article, we present the first comprehensive empirical analysis of HFS, focusing on the scope of HFS activities, the patterns of HFS crowd collaboration process, and the characteristics of HFS participant networks. A survey of HFS participants was conducted to provide an in-depth understanding of the HFS community and various factors that motivate these participants to contribute. This article also advocates a new stream of Web science and social computing research that will be important in predicting the future growth and use of the World Wide Web.

Keywords Social computing · Human flesh search · Computational sociology · Social network analysis · Cyber Movement Organizations

F.-Y. Wang (✉) · D. Zeng
The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
e-mail: feiyue@ieee.org

D. Zeng
Department of Management Information Systems, The University of Arizona, Tucson, Arizona 85721, USA

Q. Zhang
Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong, China

J. A. Hendler
Tetherless World Constellation, Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

J. Cao
The Research Center for Computational Experiments and Parallel Systems, National University of Defense Technology, Changsha 410073, China

1 Introduction

An important phenomenon happening on the World Wide Web is one in which teams of Web users get together to solve real-world problems using a combination of both online and offline resources [1]. This phenomenon, one aspect of "crowdsourcing" [2, 3], has many facets and encompasses a number of different behaviors that manifest differently across cultures and countries. Understanding such Web behaviors holds important policy and social implications; yet data-driven research is lacking. In this article we analyze one such behavior which has been occurring primarily in China, and has come to be known by the unfortunate name "human flesh search" (HFS), a literal translation from its Chinese root in which "human flesh" refers to human empowerment. There are also a number of events happened outside China that can be identified as HFS episodes, like "dog poop girl" in Korea [4] and "search for Jim Grey" in USA [1, 5], and a more recent event in which a number of British users identified a woman who had been caught on video abusing a pet.

In China, however, HFS has been increasingly employed by World Wide Web users to the point where it represents a significant amount of Web use. This HFS activity encompasses a wide variety of behaviors (Fig. 1a), with much of it now having significant real-world consequences such as identifying corrupt government officials and individuals engaging in illegal or unethical activities (e.g., animal abuse, traffic hit and runs, academic plagiarism). The first known HFS episode took place in China in 2001. However, HFS primarily gained international attention when, after the devastating Sichuan Wenchuan Earthquake in May 2008, where users cooperated through online discussion forums, chat groups and other informal Web mechanisms to help people find their missing relatives

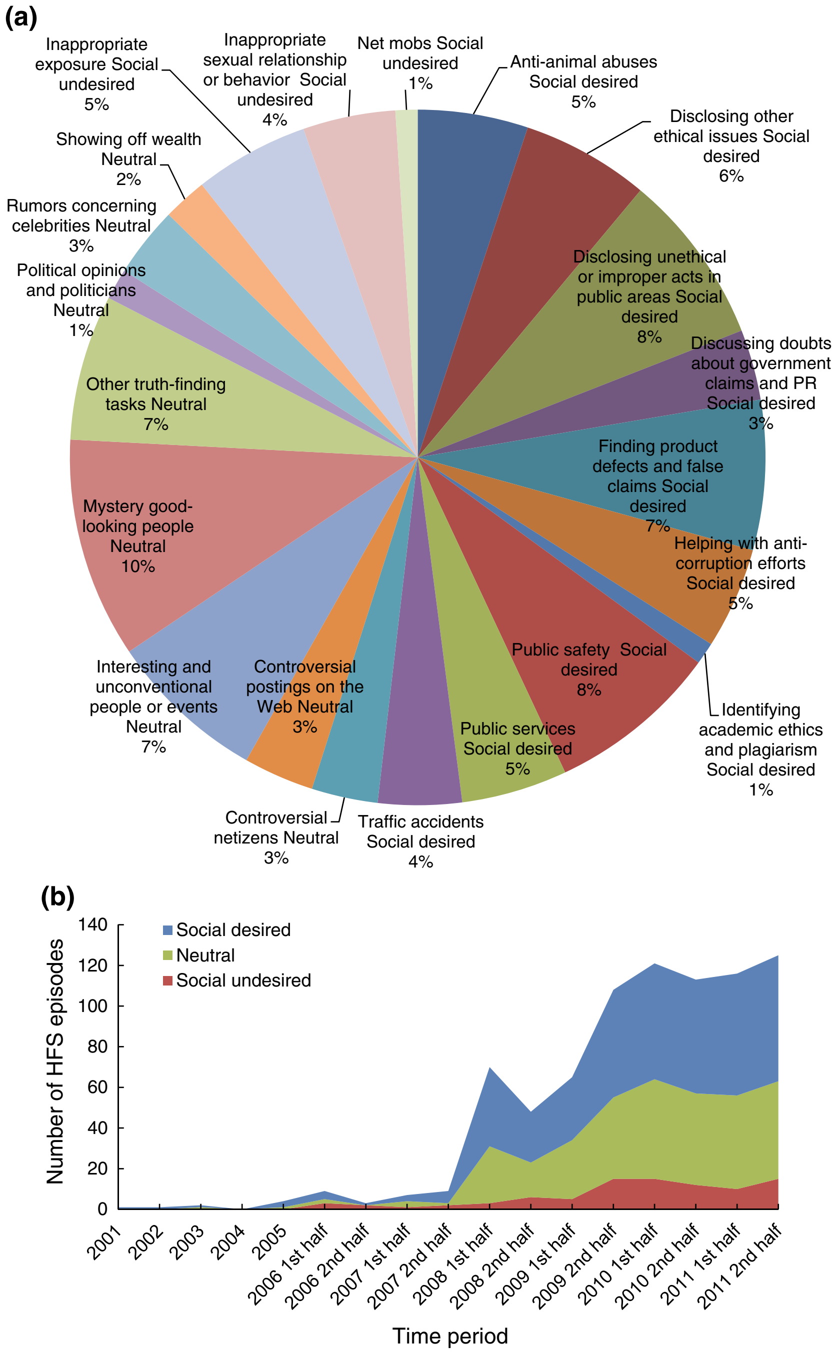

Fig. 1 **a** Types of HFS episodes, **b** evolution of HFS episodes based on social desirability

and friends. The extent of this Web use exceeded, for example, the reported use of the Web in disaster response in the US [6]. Some of the effects of HFS are also being felt outside of China and HFS episodes have fanned the interest of the general public around the global. Many non-scientific media outlets have reported on HFS. For example in October 2009, NBC aired its crime show, *LAW & ORDER*, Season 20 Episode 6, titled "human flesh search engine". In March 2010, the *New York Times Sunday Magazine* ran a report on the phenomenon focusing on several HFS episodes. In both of these, however, the media chose to highlight the negative or sensationalist uses of HFS, missing the fact that was a very small percentage of the activity, as we discuss below. In May 2012, the *International Business Times* reported on HFS describing it as a mechanism to express discontent on the Web, again missing much of the use.

In our research, we have collected the first comprehensive HFS dataset, extending one we reported in IEEE Computer in 2010 [1]. This article reports a new series of empirical analyses performed on this dataset to more fully reveal the scope of HFS activities, the patterns of the HFS crowd collaboration process, and the characteristics of HFS participant networks. We believe that HFS provides an excellent example of using the Web as a fertile ground for conducting meaningful Web science and social computing research. Such research is important in predicting the future growth and use of the World Wide Web, in advancing our understanding of societal implications of crowdsourcing, and in studying cultural aspects of World Wide Web applications.

2 Dataset and methods

To study HFS systematically, we have collected a set of online episodes dating from the inception of the approach in 2001 through December 31, 2011. Identification of these HFS episodes was done both manually, via browsing through the online forums commonly used for HFS, and by searching both traditional and social media for second-hand reporting and comments made about past HFS episodes. General Web searches were also employed to ensure the coverage of our dataset. After a particular HFS episode was identified, we first gained an in-depth understanding of its context, initiation, progression, and outcomes by going through both first-hand (e.g., postings on forums or video-sharing sites with a large number of followers) and second-hand materials (e.g., media reports) manually. We then used a Web crawler to systematically collect information from past online posts including participants' online account identifiers, these participants' IP addresses (if shown online), the full text of these posts, and the timings of replies. This allowed us to categorize the development of the behaviors and to explore the actions, both online and offline, taken by the communities involved. Using this method, we identified 802 HFS episodes occurred from 2001 to 2011, with the vast majority taking place in China (764) and the bulk occurring in the past three years (648). Since 2012, the use of *Weibo* (Chinese equivelant of Twitter) has changed the environment and platform of HFS, which made it more difficult to monitor every HFS episode. In our future research, we will build system to collect and integrate information of HFS episodes happened in and after 2012 from multiple sources.

3 An empirical study

Based on the dataset, we propose the following operational definition of HFS—"*a Web-facilitated crowd behavior aimed at accomplishing a goal-oriented task of common interest through the online sharing and disseminating information acquired from both online and offline sources without a formal platform developed specifically for the purpose*." This definition embodies the following defining characteristics of HFS: (a) Web-enabled crowd sourcing, potentially at a massive scale; (b) the need for HFS contributors to bring private, off-line information or knowledge back to the online community; and (c) online collaboration and coordination which often triggers offline actions, sometimes with serious off-line, societal consequences. Our definition is not overly restrictive as no constraints on the nature of the tasks involved (e.g., person search) are imposed. At the same time, other social computing phenomena such as crowd-sourced, Web-based, community question answering will not qualify as HFS events, which helps us better focus the unique research challenges and opportunities brought about by HFS.

Figure 1a shows a classification of HFS episodes based on their intended goals. The top category of use is finding "good looking" people from photos (10.27 %). In fact, many of those being identified go on to become Chinese Internet celebrities in their own right. The second and third largest categories are "public safety" (8.00 %) and "disclosing unethical or improper acts in public areas" (8.00 %), respectively. Since 2008, there has been an increasing use of HFS to expose government corruption, enhance public safety, identify drivers in traffic incidents, and expose problems in, or need for, public services (totaling an additional 22 %). Note that the use of HFS, contrary to the media perception, especially in the US, is increasingly being used for positive social gain and not for personal vindication.

Figure 1a presents a more aggregated view from the social desirability perspective. We have classified intended HFS goals into three groups: socially desirable, socially undesirable, and neutral. Figure 1b shows the number of HFS episodes in these three groups over the years. The mapping from the event categories in Fig. 1a to this social desirability dimension is as follows:

(1) "Socially desirable events": anti-animal abuses; disclosing unethical or improper acts in public areas, disclosing other ethical issues; public safety; helping with anti-corruption efforts; discussing doubts about government claims and public relations; traffic accidents; identifying academic ethics and plagiarism; finding product defects and false claims; public services.
(2) "Socially undesired events": inappropriate exposure; inappropriate sexual relationship or behavior; net mobs.
(3) "Neutral": controversial postings on the Web; rumors concerning celebrities; interesting and unconventional people or events; controversial Web users; mystery good-looking people; showing off wealth; political opinions and politicians; other truth-finding tasks.

According to this classification, 51.8 % of the episodes are socially desirable events, 10.7 % are socially undesirable events and the rest are neutral ones. It is evident that using HFS for societal good is becoming an increasingly important motivation for the HFS participants.

To understand how HFS works, consider the following example of a typical HFS process with the description of a major event, which we call the *Hangzhou street racing episode*. On May 7, 2009, a reckless young driver hit and killed a college student while illegally racing on the public streets in the city of Hangzhou, China. The driver's mother and friends arrived at the scene shortly. Some of the witnesses overheard the conversation between the driver's mother and his friends about how to cover up this accident. One witness also took a picture of the driver who was still smiling after the accident. This picture and the "cover-up" discussion were posted on an online forum right away, triggering the HFS. Within hours, the HFS participants were able to find that this driver's father was a well-connected and successful businessman and that the driver himself had been caught for speeding several times in the past. In addition, some of the HFS participants were able to get into this driver's personal Internet space (a password-protected personal blog shared among friends) on the largest and most popular Chinese instant messaging platform, QQ, and found that the driver kept on updating his blog after the accident. The next day the police released the initial investigation claiming that the driver was driving at the speed of 70 km/h on a street with a speed limit of 50 km/h.

The HFS participants did not believe the police report and continued their own investigation via continued HFS activity. They identified and recruited eyewitnesses and experts, trying to come up with their own estimation of the speed. The consensus, reached within five days, was that the driver was driving at a speed exceeding 80 km/h and possibly as much as 120 km/h. Both the traditional and the Internet media picked up the story from HFS findings. In response to public outcry, the police apologized officially and published a new report citing the speed at the time of the accident as between 84.1—101.2 km/h. After a court proceeding closely monitored by the public, the driver was sentenced to serve a three-year jail time and the victim's family was compensated.

Our analysis of the entire dataset (including the above episode) shows that the great majority of HFS episodes follow a common, well-structured, three-step process: (a) initiation, (b) iterative searching and sharing, and (c) post-search action and follow-up (Fig. 2). In the first step, a task with a defined goal and scope will emerge, jointly decided by a relatively small seed HFS community. In about 30 % of all HFS episodes, the initial reporting of the triggering event took place in an offline context (e.g., print-based newspapers) and the seed HFS community rebroadcast the information online to announce the event. Contrary to a common misconception that these activities are primarily incentivized by monetary gain, in only about 1.8 % of HFS episodes were cash rewards promised for information sought (HFS has been inappropriately compared in many Western media reports with the Amazon Turk (https://www.mturk.com/)) which is almost completely financially based. It is also the case that a large majority of the HFS events (91.9 %) ended in a successful conclusion (as opposed to "fading out" over time).

In the searching and sharing step, an increasingly large number of web users get exposed to HFS requests, form dynamic groups, and start intensive collaboration online, contributing their expertise in helping to do problem solving or in providing specialized knowledge not available to the wider group. A wide range of offline information- or truth-seeking activities have been observed, ranging from actually going to a crime scene in person to collect evidence, visiting landmark buildings to compare against what was shown on a photo under question, and to watching/digitizing old video tapes to match a given scene. Typically, such discussions and sharing took place in one or two online forums but for some episodes a large number of parallel discussions have been observed simultaneously with different threads of sub-topics in different forums. We also note that in about one fourth of the cases, HFS episodes benefited directly from related investigations conducted by traditional media.

In the post-search action and follow-up step, after the HFS results were publicized and, depending on the nature of the HFS task, the HFS community could engage in a range of online and offline activities. In an online context, HFS participants voiced their emotions and value judgments about the event or people. In an offline context, HFS participants made donations to help victims or to support various social causes. In some cases, there have also been negative impacts, such as enraged users harassing a suspected culprit. Interestingly, we have noticed recently that the HFS community has gotten more introspective, with the initiation phase often including an examination of the (possibly hidden, manipulative, or even malicious) purpose of the HFS requestors or originators. As abuse of the implied social contract among the HFS participants is noticed, the system essentially self-corrects, finding new ways to assure the continued integrity of the community.

4 Social network analysis of the HFS community

Careful scientific study of Web phenomena such as the HFS is an emerging area, which is not receiving the sort of attention it must from funding agencies or social Web researchers [1, 7, 8]. While there are practical reasons for this, such as the language issues and the difficulty in collecting compelling data via Web mining, studying phenomena such as HFS presents many opportunities for scientists in fields such as

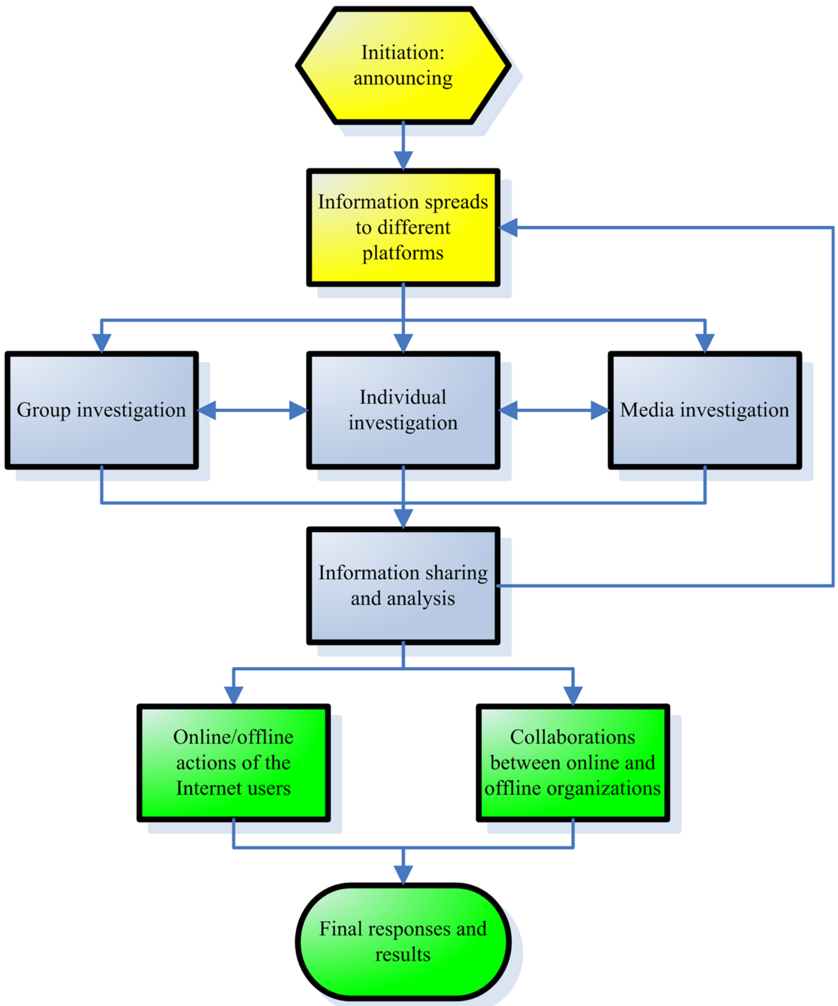

Fig. 2 (Color online) A typical HFS process

computational science [9], Web science [10], and social computing [11–13]. One such example is social network analysis of HFS groups [1, 7, 8, 14, 15].

We have conducted the social network analysis of individual HFS networks (formed from a single episode) [1] and the aggregated community of all platforms linked by inter-platform participants [16]. In order to gain a better understanding of the collaboration patterns among HFS communities, in the following we present a new analysis based on aggregated HFS networks at the platform level. We aim to answer the question whether people behave differently during HFS in different platforms, and what the difference in organizational structure in different platforms. We were able to collect associated online posts and identify the participants using their user IDs for 204 HFS episodes, out of the total 802 episodes in our dataset. Among these 204 episodes, we chose the 85 episodes that had more than 100 participants and which took place on one of the two most commonly-used HFS platforms—www.tianya.cn, the most popular HFS platform covering a wide range of general topics, and www.xitek.com, a professional online forum with a focus on photography. As Web users of different platforms have different backgrounds and behavioral patterns, we choose these two platforms to represent general-purpose and specialty HFS platforms. In order to study the collaboration patterns of the corresponding HFS community, for each platform, we constructed an aggregated HFS network by bringing together all HFS participants in one platform and building a network based on their reply-to/citation relationship. Each node in the network represents a unique user ID and an edge between two nodes represents the reply-to/citation relationship between them. The topological properties of these two HFS communities are shown in Table 1.

We observe that HFS participants on the professional *xitek* platform collaborate more and the network is much denser. We hypothesize that this difference is partially due to the scope and intensity of participants' interests. Users on *tianya* have a much wider range of backgrounds and those on *xitek* are primarily photography enthusiasts and professionals. Users on *xitek* are thus much more interested in the HFS episodes involving photography (i.e., discovering the information from a photo, finding the origin of a photo, discerning whether a photo is real or not, etc.). As such, they are more likely to participate and collaborate in those HFS episodes, resulting in a denser and more clustered collaboration network as compared to *tianya*. Interestingly, however, the larger degree-assortativity coefficient of *tianya* shows that the users from this platform act more similarly to one another in terms of their collaboration patterns. It is important to explore what caused this pattern and it is our future work.

We also analyzed the network structure of the *tianya* and *xitek* HFS communities using the bow-tie model [17], as shown in Fig. 3 and Table 2, to capture the macroscopic structure of these networks. In the bow-tie model, the network consists of five groups, including the largest strongly connected component (SCC), the nodes that only reply to the nodes in SCC (IN), the nodes that are only replied by the nodes in SCC (OUT), the nodes that either reply to OUT or are relied by IN (TENDRIL), and disconnected components (DISC). The results reinforce our conclusion made earlier, as shown by the *xitek* network having a much larger SCC, a much smaller TENDRIL, and a much smaller DISC. In particular, the majority of users on *xitek* collaborate with each other and the information flow among them is very efficient.

We also analyzed the differences in the collaboration patterns of HFS participants for different types of HFS episodes. Table 3 shows the topological properties of the HFS sub-communities on *tianya*, divided by the general HFS types—socially desirable, neutral, and socially undesirable. A participant could participate in multiple types of HFS episodes. There are 113 nodes in both socially desirable and socially undesirable sub-communities, 268 nodes in both socially desirable and neutral networks, and 63 nodes in both socially undesirable and neutral sub-communities. 10 nodes are in all three sub-communities. We have observed that the socially undesirable sub-community is denser and more clustered than the other two. The neutral network is the loosest and least clustered one among the three. In addition, the HFS participants in socially desirable and neutral sub-communities are gregarious (tending to connect to others with similar degree) as shown by the degree assortative coefficient r. The participants in the socially undesirable sub-community are not gregarious ($r < 0$), showing more heterogeneity in terms of their collaboration patterns.

5 An HFS participant survey

In order to gain an in-depth understanding of the HFS community, we conducted a survey of HFS participants with the goal of shedding light on the following critical empirical questions: (1) What are HFS participants' demographic features? (2) Why do they contribute in HFS activities? (3) What have they contributed? (4) What do they think about HFS? This survey was conducted in April 2013. We placed the survey questionnaires online and used social media platforms to invite HFS participants to share their experiences and opinions about HFS. For each participant who completed the survey, which typically took about 12 min, we offered a \$9.60 (RMB 60) online shopping voucher. To maximize the completion rate, the survey

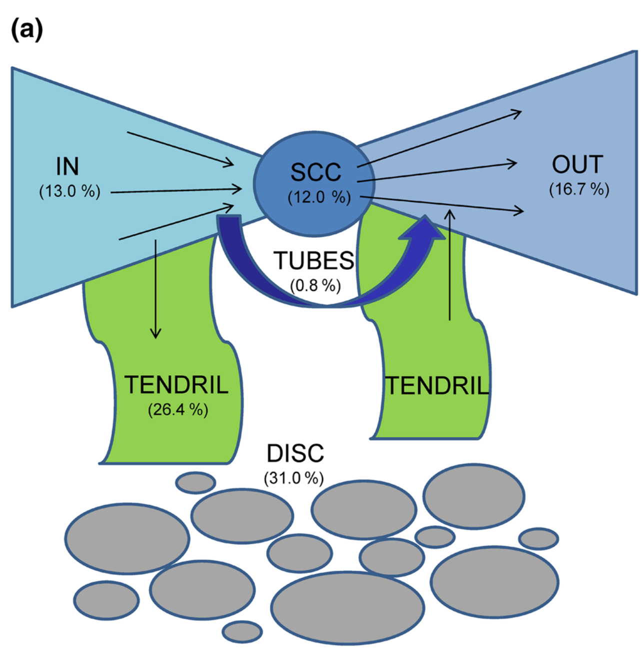

Fig. 3 (Color online) Visualization of the bow-tie structure of *tianya* (**a**) and *xitek* (**b**) HFS communities

Table 1 Topological properties of *tianya* and *xitek* HFS communities

Measure	*Tianya*	*Xitek*
N	16,706	465
L	25,396	823
Δ	0.0002	0.008
NC	2017	26
N_G (%)	11,524 (69.0 %)	414 (89.0 %)
$<d>$	2.802	3.131
C	0.027	0.037
l	8.697	5.152
D	28	17
λ_{in}	1.870	1.750
λ_{out}	1.898	1.772
r	0.11754	0.02324

N number of nodes, L number of links, Δ network density, NC number of components, N_G number of nodes in the giant component, N_G (%) the number (the portion) of nodes of the giant component (the largest weakly connected component) in the network, $<d>$ average degree, C average clustering coefficient, l average shortest path length, D network diameter, λ_{in} power of in-degree distribution, λ_{out} power of in-degree distribution, r assortativity coefficient

Table 2 The bow-tie structure of *tianya* and *xitek* HFS communities

Platform	SCC	IN	OUT	TENDRIL	TUBES	DISC
Tianya	0.120	0.130	0.167	0.264	0.008	0.310
Xitek	0.295	0.108	0.102	0.142	0.002	0.110

Table 3 Topological properties of three types of HFS sub-community

Measure	Socially desirable	Socially undesirable	Neutral
N	9756	2339	5055
L	14409	4574	6437
Δ	0.0	0.0	0.0
NC	1329	193	739
N_G (%)	6362 (65.2 %)	1735 (74.18 %)	2859 (56.56 %)
$<d>$	2.718	3.679	2.322
C	0.028	0.040	0.019
l	7.651	5.400	6.040
D	21	14	23
λ_{in}	1.825	1.443	1.858
λ_{out}	1.729	1.780	1.741
r	0.17861	−0.10822	0.06206

was conducted in Chinese. In total, we received 809 completed survey responses. The original full survey in Chinese is available online at http://www.u.arizona.edu/~qpzhang/hfssurvey/hfs-survey.pdf.

5.1 Ways of participation

Since Web users can contribute to HFS in many different ways with varying degrees of involvement and intensity, one part of our survey is aimed at identifying the range of HFS activities and the participants' activity levels (Table 4). Of those polled, nearly half have participated in one or more HFS episodes. One fourth are casual participants, who have only posted casual responses. The other participants (regular contributors) have contributed in more substantive ways. Note that although casual participants do not contribute directly, they help spread HFS information and keep the discussion in the spotlight on different HFS platforms [16]. In addition, about half of regular contributors have also performed as casual participants in certain HFS episodes. Therefore, a majority of HFS participants have been casual participants in HFS activities. This is consistent with previous findings about casual participants based on social network analysis [1, 16].

It is hard to quantify the usefulness of participants' contributions without going through manually each online post in detail in the specific context of each episode. Even then, the evaluation can be quite subjective. In our study, we analyze such contributions in a much coarse-grained fashion. We classify the type of contribution based on participants' activities during HFS. Not surprisingly, a majority of regular contributors (74.02 %) have conducted online searches. For example, they do searches online based on the existing online information and their background, and then report the results to the HFS community. This is the easiest way to contribute. The second most frequent type of contribution is knowledge contribution (24.91 %), in which participants provide clues and ideas, posting their opinions based on their own knowledge and background. It is interesting that about 6.80 % of the regular contributors have initiated one or more HFS episodes. The frequency of offline investigation and activities performed by individual participants, investigation and activities performed by groups of participants, and communications between participants either online or offline, ranges from about 7 % to about 9 %. The collaboration with real-world non-governmental organizations and government agencies are less frequent, with a rate of 4.98 % and 2.85 %, respectively. This is because (a) the fraction of HFS episodes involving collaboration with offline organizations is low [1]; (b) the number of participants who have connections or motivations to collaborate with offline organizations, especially the government, is low.

5.2 Demographics

As shown in Table 5, half of the respondents are between 24 to 30 years old. The second most frequent age range is between 18 to 30 years old. Generally

Table 4 Ways of participation

Ways of participation	Number	Fraction (%)
Type of participants		
Non-participants	435	53.77
Participant	374	46.23
Casual participant (people who have only posted casual responses like "support" and "Oh it's bad behaviors" during HFS discussions)	93	24.87
Regular contributor (people who have actually contributed to HFS instead of only posting casual responses)	281	75.13
Type of contribution		
Initiation (people who have started HFS)	19	6.80
Knowledge contribution (people who have contributed to HFS through providing clues and ideas, and posting their opinions based on their own knowledge background)	70	24.91
Online research/investigation (people who have contributed to HFS through doing online investigation)	208	74.02
Communication and discussion (people who have contributed to HFS through communicating and discussing with other participants using phone calls, text messaging, email, etc.)	26	9.25
Individual offline research/investigation (people who have contributed to HFS through doing offline research/investigation individually)	22	7.83
Collaborative offline research/investigation (people who have contributed to HFS through doing offline research/investigation as a collaborative group)	20	7.12
Individual offline activities other than research/investigation (people who have contributed to HFS through sole activities other than doing research/investigation; for example, individual protest, offline contact with the person or people involved, sole fundraising, etc.)	26	9.25
Collective offline activities other than research/investigation (people who have contributed to HFS through collective activities other than doing research/investigation; for example, group protest, group fundraising, etc.)	25	8.90
Collaboration with real-world non-governmental organizations (people who have contributed to HFS through collaborating with real-world non-governmental organizations; for example, organizations for the protection of animals)	14	4.98
Collaboration with government agencies (people who have contributed to HFS through collaborating with government agencies)	8	2.85

speaking, the distributions between non-participant, casual participant, and regular contributor follow the overall trend (Table 1) in all age groups, with non-participants most frequent, and casual participants the least frequent.

As shown in Table 5, all three types of respondents are dominated by men. In addition, men share a markedly larger fraction than women in casual participation. Although the share of women as regular contributors is only half of the share of men (48.90 %), it is still substantially larger than the findings in other online collective intelligence [18].

5.3 Incentives

A core question that has not been answered in the literature is: What incentivizes people to contribute to HFS? Table 6 summarizes our key findings. When developing the survey questionnaires, we have carefully gone through the related social science studies [19] to come up with a list of potential incentives. This list was further trimmed down to eight most likely incentives after analyzing a number of HFS episodes manually for relevance. In addition, the survey allowed the respondents to self-report what had driven them to participate in HFS activities. From the returned responses, only a handful of participants provided self-reported incentive information. As such, Table 6 tabulates findings on the eight incentives explicitly provided to the participants being surveyed.

There is one motivation standing out above all others — "Social justice, sense of community responsibility". "Belongingness to a group" is also a frequently-cited incentive (third out of eight). The second most frequent incentive is "Curiosity", which is assumed to be partially caused by the large amount of HFS episodes related to entertainment (for example, "Identifing good-looking people").

We also find that there are some participants contributing to HFS due to "Requested/Required by other people or organizations". It is interesting to align this finding with the existence of the "Internet guns-for-hire" or "water army" in the Chinese cyber-world [20, 21]. The "Internet guns-for-hire" or "water army" is a group of Internet "ghostwriters" paid to post online comments with particular content, with a certain goal. Such goals can be roughly classified into three categories: (a) Promotion of a product, a brand, people, etc.; (b) attacking the adversaries or competitors; (c) removing or mitigating negative and unfavorable posts about a product, a brand, or people. Our finding is consistent with the conjecture that the activities of the "Internet water army" can be seen in HFS.

Although money rewards and other benefits are rare in HFS [1, 16], the result shows that they do lead to more contribution, as shown by the fact that there is no one performing as a casual participant for monetary reasons. This is perhaps partly due to that it usually requires participants to find useful information to gain money reward or other benefits.

Table 5 Age and gender distribution

Age and gender	Non-participant	Casual participant	Regular contributor	Total
Age				
Younger than 18	4	1	1	6
18 to 23	137	37	88	262
24 to 30	240	52	168	460
31 to 40	43	3	19	65
Older than 40	11	0	5	16
Total	435	93	281	809
Gender				
Male	271	69	190	530
Female	164	24	91	279
Total	435	93	281	809

Table 6 Why do people contribute to HFS (%)

Incentive	Casual participants	Regular contributor
Belongingness to a group	27 (29.03)	85 (30.25)
Social justice, sense of community responsibility	69 (74.19)	197 (70.11)
Sense of personal/collective achievement	17 (18.28)	59 (21.00)
Virtual money and/or benefits	0 (0.00)	10 (3.56)
Real-world money and/or benefits	0 (0.00)	18 (6.41)
Requested/Required by other people or organizations	3 (3.23)	19 (6.76)
Curiosity	37 (39.78)	139 (49.47)
Others	0 (0.00)	5 (1.78)

6 Concluding remarks

HFS has been a unique Web phenomenon for just 10 years. In this article, we have defined HFS from a research perspective and reported on an empirical analysis of HFS based on a dataset of information on over 800 HFS episodes, aiming to understand how different types of HFS episodes over time and how HFS communities collaborate. We also conducted survey research to understand the community of greater details and the factors that affect their behaviors in HFS. Our survey research shed light on the in-depth understanding of HFS participants and people involved in the crowdsourcing systems. As we report here, most participants voluntarily contribute to HFS, without expectation of money rewards (either real-world or virtual world money). The findings indicate great potential for researchers to explore how to design a more effective and efficient crowdsourcing system, and how to better utilize this power of the crowds for social goods, solve complex task-solving problems, and even for business purposes like marketing and management.

Such analyses can generate important practical insights about modern China and its Internet use with potentially important policy implications. While in the US and Europe Web policy debates primarily focuses on privacy issues and the financial aspects of differential internet service pricing, in China HFS raises serious concerns about citizen activism and the relation between citizens and the government. In fact, the Web activism resulting from the HFS has Chinese governments at all levels experimenting with various policy alternatives related to HFS. In 2009, the Ningxia Hui Autonomous Region and the Xuzhou City Government issued decrees prohibiting online publication of personal details (such as age, income, address) of others, using privacy rules as a means to try to reduce HFS activities. At the national level, the National People's Congress Standing Committee is in the process of evaluating possibility of establishing new legislation in this area. Since March 2012, the four major microblog services in China have implemented new user registration procedurals asking for real identity, aimed at preventing the spread of online rumors and the violation of personal privacy. HFS was considered as one of the factors prompting such practice. As for HFS, the new policy is supposed to regulate the behaviors of participants and reduce the frequency of Web mobs, especially the violation of personal privacy and attacks with fake rumors. In contrast, in April 2010, Taiwan passed a law that was aimed at protecting personal information and privacy – however, this new legislation specifically legalized HFS in the context of law enforcement. The inherent inconsistencies between the needs of privacy on the Web (which is significantly less restricted in China than in the US and EU) and the socially positive nature of so much of the HFS activity are the focus of much debate in Chinese Internet policy discussions [22].

HFS provides valuable data to understand human collaboration patterns and information dissemination in social networks [23–27]. HFS also offers a natural experimental test bed and a fruitful area of study for a multiplicity of related social science disciplines [28, 29]. More importantly, the understanding of the use of the Web in China is critically important to our understanding of the future of the Web. Asia currently has more Web users than North America and Europe combined, with substantial growth yet to come. As shown in the case study of HFS reported above, the use of the Internet in China includes patterns different from those seen in the West. Studying such Web phenomena originated in the East is important to our ability to understand the cultural differences of the use of the Internet and how international policies which might be well-intentioned in the West, such as privacy legislation, can have adverse effects in other cultures.

In the past decade, there have been many successful applications of social computing and collective intelligence including HFS, crowdsourcing [2, 30], social search [31], collective decision making[32, 33], collective design and manufacturing [34], and more. These revolutionary applications on the Web have shed light on how the power of collective intelligence have been dramatically changing the way that people communicate, collaborate, research, innovate, and produce in the age of connected world, which lead to an emerging part of the Web Science research field which we refer to as Cyber Movement Organizations (CMO) [35]. Since the social structures and organizations have been changed dramatically as a result of those applications, in-depth research on Web phenomena like HFS from the perspective of CMO has the potential to help us further understand and architect the future of Web, and the society. In addition, how to utilize similar crowd power to facilitate collective intelligence and social machines [36] is of great interests to social computing communities and is part of our on-going research.

Acknowledgement The authors thank Guanpi Lai for his help in collecting data during this research. This work was supported in part by the National Natural Science Foundation of China (90924302, 91024030, 71025001, 70890084, and 60921061), the US Defense Advanced Research Projects through two seedling grants to Rensselaer Polytechnic Institute, and the US National Science Foundation support for EAGER (IIS-1143585).

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Wang F-Y, Zeng D, Hendler JA et al (2010) A study of the human flesh search engine: crowd-powered expansion of online knowledge. Computer 43:45–53
2. Howe J (2009) Crowdsourcing: why the power of the crowd is driving the future of business. Three Rivers Press, New York
3. Doan A, Ramakrishnan R, Halevy AY (2011) Crowdsourcing systems on the world-wide web. Commun ACM 54:86–96
4. Kietzmann J, Angell I (2010) Panopticon revisited. Commun ACM 53:135–138
5. Hellerstein JM, Tennenhouse DL (2011) Searching for Jim Gray: a technical overview. Commun ACM 54:77–87
6. Palen L, Hiltz SR, Liu SB (2007) Online forums supporting grassroots participation in emergency preparedness and response. Commun ACM 50:54–58
7. Wang B, Hou B, Yao Y et al (2009) Human flesh search model incorporating network expansion and GOSSIP with feedback. In: Turner SJ, Roberts D, Cai W (eds) IEEE/ACM International Symposium on Distributed Simulation and Real-Time Applications. 13th IEEE/ACM Symposium on Distributed Simulation and Real-Time Applications (DS-RT 2009), Singapore, October 2009. IEEE Computer Soc, DC, pp 82–88
8. Ahn Y, Bagrow J, Lehmann S (2010) Link communities reveal multiscale complexity in networks. Nature 466:761–764
9. Bonabeau E, Dorigo M, Theraulaz G (2000) Inspiration for optimization from social insect behaviour. Nature 406:39–42
10. Hendler J, Shadbolt N, Hall W et al (2008) Web science: an interdisciplinary approach to understanding the web. Commun ACM 51:60–69
11. Wang F-Y, Carley KM, Zeng D et al (2007) Social computing: from social informatics to social intelligence. IEEE Intell Syst 22:79–83
12. Zeng D, Wang F-Y, Carley KM (2007) Guest editors' introduction: social computing. IEEE Intell Syst 22:20–22
13. Wang F-Y (2007) Toward a paradigm shift in social computing: the ACP approach. IEEE Intell Syst 22:65–67
14. Barabási A (2009) Scale-free networks: a decade and beyond. Science 325:412–413
15. Clauset A, Moore C, Newman M (2008) Hierarchical structure and the prediction of missing links in networks. Nature 453:98–101
16. Zhang Q, Wang F-Y, Zeng D et al (2012) Understanding crowd-powered search groups: a social network perspective. PLoS One 7:e39749
17. Broder A, Kumar R, Maghoul F et al (2000) Graph structure in the web. Comput Netw 33:309–320
18. Glott R, Schmidt P, Ghosh R (2010) Wikipedia survey—overview of results. United Nations University: Collaborative Creativity Group
19. Zhang Q (2012) Analyzing cyber-enabled social movement organizations: a case study with crowd-powered search. University of Arizona, Tuscan
20. Chen C, Wu K, Srinivasan V et al (2011) Battling the internet water army: detection of hidden paid posters. arXiv preprint arXiv:11114297
21. Wang G, Wilson C, Zhao X et al (2012) Serf and turf: crowdturfing for fun and profit. In: Proceedings of the 21st international conference on World Wide Web. ACM, New York, pp 679–688
22. Liang B, Lu H (2010) Internet development, censorship, and cyber crimes in China. J Contemp Crim Justice 26:103–120
23. Eubank S, Guclu H, Anil Kumar VS et al (2004) Modelling disease outbreaks in realistic urban social networks. Nature 429:180–184
24. Kleinberg J (2007) Computing: the wireless epidemic. Nature 449:287–288
25. Gonzalez MC, Hidalgo CA, Barabasi AL (2008) Understanding individual human mobility patterns. Nature 453:779–782
26. Song C, Koren T, Wang P et al (2010) Modelling the scaling properties of human mobility. Nat Phys 6:818–823
27. Song C, Qu Z, Blumm N et al (2010) Limits of predictability in human mobility. Science 327:1018–1021
28. Lazer D, Pentland A, Adamic L et al (2009) Computational social science. Science 323:721–723
29. Watts DJ (2007) A twenty-first century science. Nature 445:489
30. Pickard G, Pan W, Rahwan I et al (2011) Time-critical social mobilization. Science 334:509–512
31. Chi EH (2009) Information seeking can be social. Computer 42:42–46
32. Introne J, Laubacher R, Malone T (2011) Enabling open development methodologies in climate change assessment modeling. IEEE Softw 28:56–61
33. Introne J, Laubacher R, Olson G et al (2011) The climate CoLab: large scale model-based collaborative planning. Collab Technol Syst 40–47
34. Yu L, Nickerson J, Sakamoto Y (2012) Collective creativity: where we are and where we might go. Collect Intell (CI'12)
35. Wang F-Y (2011) Study on cyber-enabled social movement organizations based on social computing and parallel systems. J Univ Shanghai Sci Technol 33:8–17 (in Chinese)
36. Hendler J, Berners-Lee T (2010) From the semantic web to social machines: a research challenge for AI on the World Wide Web. Artif Intell 174:156–161